DOCUMENT RESUME

ED 383 757 TM 023 287

AUTHOR Griffin, Marlynn M.; And Others
TITLE The Effects of Cooperative Assessment on Goals, Perceived Ability, Self-Regulation and Achievement.
PUB DATE Apr 95
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Ability; *Academic Achievement; Cognitive Processes; Cooperative Learning; Higher Education; *Objectives; Pretests Posttests; Psychology; *Student Attitudes; Student Motivation; Test Construction; *Undergraduate Students
IDENTIFIERS *Cooperative Assessment; *Self Regulated Learning

ABSTRACT

Cooperative assessment was investigated in a classroom setting, examining achievement outcomes, effects on motivation, and student perceptions of the cooperative assessment process. Eighty-four undergraduate psychology students participated in this nonequivalent control group study design. It was hypothesized that students taking tests using a cooperative assessment procedure would perform significantly better on a posttest of educational psychology course concepts than would students completing tests in a traditional format. Effects of the treatment on goal orientation, perceived ability, self-regulation, and depth of processing were examined. Analysis of covariance indicated that there were no significant differences between the groups on the posttest, and that the hypothesis was not supported. There were also no differences between groups on measures of goal orientation, perceived ability, and depth of processing. Student reactions to the cooperative assessment procedure were overwhelmingly positive. Students enjoyed taking tests in groups, and felt that they learned more through this process as they discussed and debated the responses to the test items. Two tables illustrate the discussion, and two appendixes provide supplemental information. (Contains 24 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Marlynn Griffin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

The Effects of Cooperative Assessment on Goals, Perceived Ability, Self-regulation and Achievement

ED 383 757

Marlynn M. Griffin
Bryan W. Griffin
Robert W. Warkentin
Georgia Southern University

Gwendolyn P. Quinn
Marcy P. Driscoll
Florida State University

R. R. McCown
Duquesne University

Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, April 18 - 22, 1995.

TM023287

BEST COPY AVAILABLE

The Effect of Cooperative Assessment on Goals, Perceived Ability, Self-regulation, and Achievement

Marlynn M. Griffin, Bryan W. Griffin, Robert W. Warkentin, Georgia Southern University
Gwendolyn. P. Quinn, Marcy P. Driscoll, Florida State University
R. R. McCown, Duquesne University

Abstract

Amongst the plethora of cooperative learning studies, several investigations of collaborative, cooperative, or group assessment have appeared. These studies have investigated cooperative assessment in laboratory conditions (Lambiotte, et al., 1987), and in classroom settings (Bilsky-Torna, 1993; Webb, 1993; McCown, 1992), and have examined the effects of cooperative assessment on learning (Lambiotte, et al., 1987; Bilsky-Torna, 1993; Webb, 1993; R. R. McCown, personal communication, April 5, 1993) and group process on assessment outcomes (Webb, 1993). In this study, we refined McCown's methodology and investigated cooperative assessment in a classroom setting, examining achievement outcomes, effects on motivation, and student perceptions of the cooperative assessment process.

Eighty-four undergraduate educational psychology students participated in this non-equivalent control group design study. It was hypothesized that students taking tests using a cooperative assessment procedure would perform significantly better on a posttest of educational psychology course concepts than would students completing tests in a traditional format. In addition, the effect of the treatment on goal orientation, perceived ability, self-regulation, and depth of processing was examined. The cooperative assessment group completed exams individually and then in groups; student exam grades were a combination of individual and group exam scores. The traditional assessment group took exams individually.

Analysis of covariance indicated that there were no significant differences between the groups on a posttest of educational psychology concepts, thus the hypothesis was not supported. There were also no differences between groups on measures of goal orientation, perceived ability, and depth of processing, though there were statistically significant differences on self-regulation. Student reactions to the cooperative assessment process, however, were overwhelmingly positive. These data indicated that although there were no statistically significant differences in achievement between treatment groups, students enjoyed taking tests in groups and felt that they learned more through this process as they discussed and debated the responses to test items.

Cooperative learning has been thoroughly documented as an effective learning tool and teaching strategy (Johnson & Johnson, 1989; Johnson, Johnson, & Smith, 1990; Slavin, 1991). Cooperative learning, students learning from and with each other, helps to develop social skills, establish professional working skills, builds a sense of community within a classroom, and enhances student achievement, self-esteem, and attitudes toward school at all grade levels, including in college classrooms (Johnson, Johnson, & Smith, 1990; Slavin, 1991; Wynne, 1983). One explanation for these effects is that cooperative learning techniques augment the extent to which content is actively processed by students and offers participants the opportunity for discussion and negotiation which may lead to higher-level reasoning and the development of strategies which facilitate these processes (Gabbert, Johnson, & Johnson, 1987; Johnson et al., 1990; Nystrand, 1986). The social support which often arises from a cooperative task has been shown to aid students in persisting on a challenging task, reducing frustration, increasing autonomy, and contributing to academic and career aspirations (Gabbert, et al., 1987; Sarason, Sarason, & Linder, 1983).

Cooperative techniques require students "to explain what they are learning to each other, learn each other's point of view, give and receive help from classmates, and help each other dig below the superficial level of understanding the material they are learning" (Johnson, et al., 1990, p. 11). While such "digging" and sharing of viewpoints can and does occur during cooperative in-class practice activities which take place as part of instruction, it is also likely that cooperative learning groups will negotiate under test conditions. Many believe that this negotiation of understanding is essential for knowledge construction (Duffy & Bednar, 1991; Kember & Murphy, 1990; Vygotsky, 1978). Brown (1989) argued that learning is "about the making of meaning, not just the receiving of it. Thoughtfulness is a constructive, not a passive, undertaking" (p. 32).

Several studies have begun to look at the benefits of an extension of the cooperative learning process - cooperative assessment. Lambiotte et al. (1987) found, under laboratory conditions with college students, that cooperative test-taking led to increased quantity of recall in reading and comprehension. Singer (1991) looked at the efficacy of cooperative testing with junior high students in a pre-algebra class. His study revealed an increase in test scores using cooperative test-taking, as well as an expressed preference by the students for this form of test taking. R. R. McCown (personal communication, April 5, 1993) reported similar findings in an unpublished pilot study of cooperative assessment.

Other studies have investigated the effects of group or collaborative assessment on achievement and social process. Bilsky-Torna (1993) found that assigning grades based on group responses to an English quiz increased student motivation and the quantity of communication between students and teacher during the test. This study also reported disadvantages to cooperative testing such as group instability, noisy classrooms, "stifling" of academically stronger students, and weaker students riding on the coattails of stronger students during the group quizzes.

Webb (1993) analyzed the relationship between achievement scores obtained during small-group assessment tasks and individual assessment tasks. In Webb's study, students solved mathematics operations on decimal numbers in collaborative small groups for a 50-minute class period. Two weeks later, following a review session, students examined a similar problem without collaborating with other students. Students consistently performed better in the group assessment despite prior ability whereas student performance on the individual assessment was accurately predicted by both ability and behavior within the small group setting.

Proponents of authentic assessment have called for assessment to be more than an end product. Rather, assessment should be part of the learning process (c. f., Shavelson, Baxter, & Pine, 1992; Shepard, 1989; Wiggins, 1989). Many educators have had the experience of returning a test to the class and attempting to help students learn from their mistakes, only to find that many of the students are far less interested in learning from their mistakes than they are in trying to rationalize their incorrect responses in the hopes of earning extra points. Cooperative assessment places the ownership for learning in the hands of the students and offers opportunities for the negotiation of understanding. This negotiation can be a valuable learning tool and can encourage

students to think about what they have learned and are learning as they discuss the assessment with their peers.

In cooperative assessment, students complete unit examinations individually and immediately afterwards in small groups. Since cooperative assessment yields grades that are equal to or higher than an individual test score, never lower (Griffin, et al., 1994), a possible negative outcome of the procedure is that students may look upon the group assessment portion of the process as a way to earn "free" points and subsequently decrease the amount of emphasis they place on preparing for an examination and monitoring their study strategies (self-regulation), and on the value of the assessment as a learning task rather than an opportunity merely to earn more points (goal orientation). This possible side-effect must be investigated before any assumptions about the efficacy of cooperative assessment can be made. Pintrich and DeGroot (1990) included three components in a working definition of self-regulation: (a) metacognitive strategies used by students in planning and monitoring learning (Brown, Bransford, Campione, & Ferrara, 1983; Corno, 1986); (b) control and management of effort for classroom tasks (Corno, 1986); and (c) type of cognitive strategies students use in learning and remembering (Corno & Mandinach, 1983). Learning goals have been described as goals which are pursued with an emphasis placed on becoming more proficient at a task, while performance goals are typically pursued for the purpose of looking good in front of others (Elliott & Dweck, 1988). Greene and Miller (1993) developed the Motivation and Strategy Use Survey to examine self-regulation and goal orientation as well as perceived ability and depth of processing.

This study, then, investigated the effects of cooperative assessment on individual achievement, goal orientation, depth of processing, self-regulation, and perceived ability in an educational psychology class for preservice teachers. The design differs from that utilized by Lambiotte (1987) in that it occurs in a classroom context rather than a laboratory context, and is thus more ecologically valid. It also differs from those utilized by Webb (1993) and Bilsky-Torna (1993) in that students take tests first individually and then in groups to ensure individual accountability as well as group rewards. We have employed the design utilized by Griffin, et al., which was developed by R. R. McCown (personal communication, April 5, 1993) who found positive effects on student grades and student attitudes using this design.

Method

Subjects and Design

Eighty-five students enrolled in four sections of educational psychology, taught by the first author, participated in the study. All students were education majors with a variety of areas of emphasis (e.g., early childhood, middle grades, physical education, etc.). There were 16 males and 31 females in the treatment group, and 9 males and 29 females in the control group. 9 students in the experimental and 4 students in the control group were Black, while the remainder were White.

A non-equivalent control group design was utilized, with treatments randomly assigned to groups. Scores from a pretest (discussed below) and self-reported GPAs were obtained and analyzed to determine if groups differed significantly in these areas. The groups were not statistically different on GPA $F(1, 84) = 1.46$; $p = .23$ though initial differences did exist on pretest performance, control $M = 13.17$, $SD = 2.50$, treatment $M = 11.91$, $SD = 3.14$, $F(1, 84) = 4.2$; $p = .043$. Analysis of covariance was utilized in later analyses in an effort to provide some statistical control for these initial differences in pretest scores.

The independent variable was testing condition in which participants either took tests first individually and then as part of a group (cooperative assessment) or completed tests on an individual basis only (individual assessment). Students in the treatment group could add up to ten points to their individual grade if performance on the group assessment exceeded their individual performance, but would not lose any points if individual performance surpassed group performance. Cooperative scores (i.e., the student's combined individual and group assessment scores) could not exceed group scores. For example, suppose Group A obtained a 96 on their group exam. Student A1 obtained an 82 on her individual exam, and is thus eligible for all 10 group points, bringing her cooperative assessment score to 92. Student A2 scored 88 on the

individual exam, and is thus eligible for only 8 group points so that she does not exceed the group score, giving her a 96 on her exam. A score of 97 was obtained by Student A3 on the individual assessment, so she does not receive any group points and she keeps her grade of 97. Using this scoring procedure, we attempted to provide for both individual accountability and group incentives. Since students could add only up to 10 points to their individual score, they had to do some advance preparation for the exam to obtain a passing grade. By allowing students to earn extra points to add to their individual grades, incentive to work together as part of a group was also established as it was in everyone's best interest to obtain the highest group score possible to maximize the likelihood that students would earn group points and thus increase their exam grade.

The dependent variables were achievement, as measured on a posttest of educational psychology course content and perceived ability, goal orientation, and self-regulation, as measured by the Motivation and Strategy Use Survey (Greene and Miller, 1993). Survey data from the cooperative assessment group evaluating the cooperative assessment strategy and follow up interviews with representative students in the treatment condition were also collected.

Materials

Pretest. The pretest consisted of 30 multiple choice items, 10 from each of the first three units of instruction for the introductory educational psychology course. The topics addressed on the pretest were operant conditioning and information processing (unit 1), observational learning, motivation, and outcome decisions (unit 2), and instructional models, instructional tactics, and classroom management (unit 3). The items on the pretest were primarily application type multiple choice items, and all were matched to unit objectives to ensure content validity.

Posttest. The posttest was comprised of the same 30 items which appeared on the pretest and addressed the **first** three units of instruction. The posttest was administered in three sections, 10 questions on each of the **last** three unit exams. Thus, the last three exams were partially cumulative. The 10 cumulative items from each of the last three exams were added together to comprise the posttest, and student posttest score was defined as the number correct of these 30 items. This partially cumulative approach to testing was designed to provide a measure of the testing condition effects on individual learning. Using this strategy, students completed test items from unit 1 on the individual portion of the test. Next, students in the cooperative assessment condition discussed the same items from unit 1 in their groups. Then, on the second unit test, students completed the 10 posttest items from unit 1 (as well as the 30 items from unit 2) on an individual basis prior to discussing test 2 with their groups. The posttest scores were taken from the **first** administration of the unit 2, 3, and 4 exams to measure the impact of the treatment on individual performance.

Motivation and Strategy Use Survey. This 57 item survey was administered to students immediately prior to completing the fourth exam. One item was omitted from our analysis because many students found it ambiguous. This item read, "Attending class was only helpful when something important for the test was covered."

Greene and Miller (1993) completed a factor analysis of the survey and found items loading on six factors. See Table 1 for a listing of the factors, the number of items per factor, and the internal consistency measures of each factor for the original Greene and Miller study as well as for our analysis.

Table 1

Factor (# items)	Greene alpha	Griffin alpha
learning goal (4)	.86	.74
performance goal (4)	.58	.74
perceived ability (8)	.74	.69
self-regulation (17)	.87	.77
deep processing (11)	.83	.83
shallow processing (12)	.73	.74

All items were measured on a five-point Likert scale, ranging from Strongly Disagree (1) to Strongly Agree (5) and reverse-scored items were entered into the analysis accordingly.

Group Assessment Survey. A 12 item survey was designed to tap student opinions and perceptions of the cooperative assessment process. These twelve items examined group process and student perceptions of the testing situation (See Appendix A).

Procedures

On the first day of the quarter all students completed the 30 item pretest of educational psychology concepts. Students also completed a questionnaire requesting demographic information which was used to assign students to heterogeneous groups.

Students in all classes were assigned to cooperative groups of 4-5 members, which were mixed on the basis of gender, race, ability, and major. In terms of ability, each group consisted of a student with a high GPA (3.5-4.0), two students with average GPAs (2.5-3.49), and one student with a low GPA (below 2.49). Males and Black students were divided among the groups, and, to the extent possible after mixing the groups on all other variables, each group contained at least one student who was not an early childhood education major (the predominant major in all classes). Students in all classes completed other types of group activities as well as cooperative assessment. Students worked in their groups to complete in-class non-graded activities designed to facilitate acquisition of course concepts. Each group also worked to cooperatively complete a series of article critique papers concerning issues in educational psychology. These papers were graded and all group members received the same grade on each of the papers (4-5 papers, depending on the number of people in the group). In addition, the students in the cooperative assessment group completed the second administration of each unit exam working cooperatively with their group members.

With the exception of the difference in testing procedures, all sections of the course were taught in the same manner. The two control group sections took place, one section each, during the summer and fall quarters of 1994. The experimental group sections were taught during the winter quarter of 1995. Care was taken by the instructor to use similar examples, to complete the same activities, and to cover the same amount of material per period in all classes. There were occasions, however, on which one group discussed something to a different depth or from a different angle than another group. Although this does not allow for strict control of the teaching conditions, more effort to deter these discussions was not made because the instructor was unwilling to interfere with the quality of the instruction in order to implement strict experimental controls. Class sessions were typically a mixture of lecture, discussion, generation of examples, examination of practical applications of course content, and group activities.

At the end of each of the four units students completed a unit examination. Each exam was comprised of 40 multiple choice questions (primarily higher level items) and 2 essays items. Students in the experimental group were allowed to leave the room after completing the individual portion of the exam and were instructed to return at a time designated by the instructor. Students in the experimental group worked together to complete the test a second time after returning to class. Each group was permitted to submit only one set of responses per group, so students were forced to come to consensus on the responses to the test items.

Exams were returned to students within two class periods for all sections of the course. During the period in which exams were returned, the instructor placed an answer key on the overhead projector so students could check the accuracy of the machine scoring. Copies of the exam were not distributed to students, but copies were available in the instructor's office for students' perusal. Very few students in either treatment condition chose to pursue this option.

Immediately prior to completing the final examination all students completed the Motivation and Strategy Use Survey. Students in the experimental group also completed the Group Assessment Survey and submitted it on the day of the final examination. Following the final examination, students were informed that they had been participating in an experiment throughout the quarter. They were briefly told of the purpose of the study and assured that individual data would be kept confidential. At this point, students were also told that they could request that their data be withdrawn from the study if they chose not to participate. Students could exercise this

option immediately or after grades were reported, in case they feared retribution for withdrawing their data from the study. No student chose to withdraw his or her results from the data pool.

Toward the end of the quarter in which data were collected for the treatment group, eight students from the cooperative assessment condition were asked to come in and participate in individual interviews with the third investigator. The third investigator was selected to conduct the interviews, rather than the first, in case students had any concerns to voice about the procedure. We anticipated that students would be less likely to offer criticisms of cooperative assessment to the instructor of the class in which they were completing the cooperative assessment procedure. Stratified random sampling was used to select the students; stratifications were based on self-reported GPA using the same categories used to assign students to cooperative groups. Initially, nine students agreed to complete interviews, but one student from the higher GPA group later was unable to attend the interview. The interviews were audiotape recorded (with the permission of the student). Each student was asked the same series of questions, but responses were explored in more depth as needed.

Results

Posttest

All data were analyzed at the alpha = .05 level. Posttest data were analyzed using ANCOVA, with pretest scores entered as a covariate. No significant differences were found between the groups on the posttest, control group $M = 21.03$, $SD = 3.29$, treatment group $M = 21.49$, $SD = 3.66$, $F(1, 80) = 2.18$; $p = .14$. The test for an interaction between pretest and treatment also yielded statistically insignificant differences, $F(1, 80) = 1.32$; $p = .25$. The effect of pretest scores $F(1, 80) = 4.31$; $p = .041$ and GPA was significant, $F(1, 80) = 20.40$; $p = .000$, indicating a positive relationship between pretest scores and posttest performance and GPA and posttest performance. Thus, the hypothesis of this study was not supported.

Motivation and Strategy Use Survey

The items for each of the six factors were aggregated within those factors and analysis of variance was conducted to determine if differences between the groups existed on any of the six factors. The results of these analyses are presented in Table 2.

Table 2

Factor	df	p-value
learning goal	$F(1, 82) = 1.24$	.27
performance goal	$F(1, 82) = .02$	.89
perceived ability	$F(1, 82) = .65$	.42
self-regulation	$F(1, 82) = 6.54$	.01
deep processing	$F(1, 82) = .04$	.85
shallow processing	$F(1, 82) = .04$	.85

The groups were not different at any traditional levels of significance on any of the factors except self-regulation. For this factor, the control group exhibited a higher mean than the treatment group.

Survey

Analysis of survey data indicated that student perceptions toward cooperative assessment were quite positive. Mean scores, modes, and standard deviations for the Group Assessment Survey are presented in Appendix A (note that items 3, 9 and 12 are scored in reverse). These results indicate that students tended to discuss test items with group members, prepared about the same amount for exams in this course (despite the fact they would receive extra points from group collaboration) as they did for exams in other courses, and felt that taking tests in groups was somewhat beneficial to their grades. Students also believed that they were learning more about the course content by discussing the exams with their groups.

Interview

The last source of data to be presented, interview data, was compatible with the outcomes of the survey. The questions asked during the interview fell into two major categories, affective and cognitive/metacognitive outcomes. Students indicated positive responses to most questions in both categories. That is, they reported that they were interested in learning, they valued the task, and they were determined to find the best answers and best supporting rationales for these answers.

In sum, then, the outcomes on the posttest do not support the hypothesis of this study. Student survey and interview responses to the cooperative assessment strategy, however, were overwhelmingly positive.

Discussion

Overview and Posttest Outcomes

This study examined the effect of cooperative assessment on the acquisition of educational psychology course content. It was predicted that students in the cooperative assessment group would outperform students in the individual assessment group on a posttest of educational psychology. This hypothesis was not supported. Certainly the first explanation to be considered regarding this finding is that cooperative assessment does not improve learning over an individual assessment format, as the results seem to indicate. Student perceptions of the procedure, however, appear to contradict this finding. As will be discussed in greater detail, students reacted positively to the treatment and felt that they learned more from this method of testing than from traditional methods. Still, it must be noted, student perceptions are not always accurate indicators of achievement gains.

Survey

The results of the survey administered at the end of the quarter indicated a generally positive regard by students for the cooperative assessment procedure. Survey items used a 7-point Likert scale with higher numbers indicating more positive responses except for the reverse scored items, 3, 9, and 12 (see Appendix A). Three major areas were emphasized in the survey questions: group testing process, study habits, and benefits of cooperative assessment.

Items 2, 4, 9, and 10 addressed the processes through which the groups went as they completed tests together. Note that item 9 is reverse scored, so that the lower mean corresponds with positive responses on items 2, 4, and 10. The survey indicates that students did tend to discuss the responses to test items more than they tended to rely on one person to provide a correct response and that all group members offered input to the discussion of individual items. Thus, it appears that students negotiated the responses to test items and attempted to reach a mutual understanding about the content. This finding is consistent with the results of the interview which indicate that for some items there was group discussion of the correct response, but for other items, group consensus, rather than discussion, was the norm. Observations of the group testing phase by the instructor confirm that students did not discuss all items, but primarily discussed those items for which group members had selected a variety of responses. For these items, however, lively debates often ensued with different members trying to convince others of the accuracy of their individual response.

Preparation for unit exams was addressed in items 1, 5, and 6. It seems quite reasonable that students might study less for an exam on which they have the opportunity to improve their grades through group testing than in a traditional testing situation in which the grade they earn is solely the result of their own efforts. This possibility was, in fact, what prompted us to allow the group exam to count only a maximum of 10 points toward a student's exam grade. If students can earn 10 points, there is incentive to work with the group to earn the highest grade possible. There is also incentive, however, to prepare for the individual test since one must do reasonably well in order to pass the exam even if all 10 points were earned. Students indicated that they and other members of their group were well-prepared for the exams (item 1). Survey results (item 5) indicate that students studied about the same amount as they usually do, despite knowing they had a group exam to complete. Of slightly more consequence to the students' preparation was the

difficulty level of the exams (item 6); students indicated that they prepared somewhat more than usual due to the difficulty of the exams.

Last, the survey examined the potential benefits of the cooperative assessment method. As in interviews, students focused on both the benefits of cooperative assessment to their course grades as well as to their understanding of course concepts. Note that items 8 and 11 (intrinsic, learning goal) have higher means and lower standard deviations than item 7 (extrinsic goal). Question 12 also looked at students' goal orientation, and seems to indicate that students leaned more toward a learning goal orientation. Note that the lower mean on this reverse scored item indicates that students were concerned with receiving feedback from the group assessment, but were somewhat more concerned with the benefits of learning from their group. Overall, then, students believed that the cooperative assessment procedure did benefit their understanding of course material. This finding is another plus for using cooperative assessment - students perceive they are learning through this method as they work in groups to complete testing tasks. Group discussion of exams takes the responsibility for dispensing knowledge to the students out of the hands of the teacher and places it into the hands and minds of the students, encouraging them to be responsible for their own acquisition and construction of knowledge.

Interview

The interviews focused primarily on issues related to affective outcomes (goals, social interactions of group members, self-efficacy) and cognitive/metacognitive outcomes (planning, monitoring and regulating group testing, evaluating, form of discussion) (see Appendix B). We will first discuss the results pertaining to the affective outcome questions, followed by responses to cognition/metacognition questions.

When asked what their goal was during the group test, all students mentioned achieving a better grade or getting more points on the test. This seems to indicate that students were externally motivated. However, students also indicated that they wanted to learn and understand the information on the exams, indicating some degree of internal motivation as well. They valued the cooperative assessment task as being worthwhile and wanted each person in the group to participate. Students indicated that they evaluated each others' responses as possible ways to understand the information. Perhaps then, students were motivated to participate in group discussions by both internal and external forces.

The procedure used by some of the groups to decide which items to discuss seemed to preclude many (perhaps too many) items from being debated and discussed. In some groups, if there was no disagreement about the correct response to an item, the item was not even read aloud. This appeared to limit the number of items, and therefore the amount of content, discussed. Perhaps, then, problems did not lie in the level or quality of the discourse, but the breadth of the discourse. As for the items which were discussed, we know little about them. Students were not asked to indicate which items they discussed nor to what extent they were discussed. Items may have been discussed because they were vague or ambiguous, not because they were particularly important. Perhaps, then, these items constrained students' attention to details or unimportant information. If students found it necessary to debate all test items maybe the strength of the cooperative assessment procedure would be increased.

Student self-efficacy was not challenged as much as it could have been. Some students did not feel completely challenged because they did not have to discuss every item; they could skip many items. Most students said that they were challenged when they were debating and discussing items, particularly if they were certain their response was correct and were trying to defend that response. Some students also mentioned that they did not have to risk much to participate in this exercise; after all, their individual grade could not go down, it could only be improved. Thus, there is little to lose if the group does not improve. Perhaps self-efficacy could be better challenged (and subsequently improved) if the frequency of the discussions increased.

Cooperative assessment did not seem to change students' regular test preparation tactics. That is, they did not decrease their amount of studying as a function of knowing they could rely on the group for extra points and increased grades. They also did not, however, modify their studying to accommodate the different testing procedures required by the group test. That is, they

assumed that their regular test preparation tactics were appropriate for group testing. It is possible that their regular strategies were perfectly adequate for cooperative assessment. It is also possible, however, that students could have benefited more from the procedure if they had modified their study habits to accommodate group discussion as a function of cooperative assessment.

Students perceived the cooperative assessment activity as a problem solving task. They focused on explaining why their answers were correct or incorrect and tried to clear up confusing issues. They also evaluated each other's reasoning about information and tried to determine the best answers and rationales for the answers. Furthermore, they discussed different perspectives on particular items and valued comparing different viewpoints. Students indicated that they were actively monitoring their peers' discussion when there was disagreement, especially when there was an "even split" with equal numbers of group members arguing for two different responses. Many diverse examples were generated during group discussion, indicating that students were clarifying and re-examining their knowledge of the information through interaction. However, all of this discourse activity was limited to the few items which were actually discussed. The question remains, then, how can students be encouraged to discuss all of the test items?

One possible, and obvious, solution to this problem is to have students work on the tests as a group from the outset - that is, eliminate the individual portion of the assessment. While this might encourage more discussion, it also eliminates the provision made for individual accountability and may encourage more "free-riders." Another possible solution is to generate test items of a different type and/or difficult enough across all items to encourage more discussion. A third potential solution, and one which will be investigated in future studies, is to generate different items for the individual and group assessments. By taking this measure, individual accountability could be retained while students would be more likely to, minimally, read all items aloud, and, optimally, discuss all of the items because they had not seen any of them previously.

References

Bilsky-Torna, Z. (1993). "Son of S.T.A.D." - A comparison of group quiz and individual test scores. English Teachers Journal, 46, 35-43.

Brown, A. L., Bransford, J. D., Campione, J. C., & Ferrara, R. A. (1983). Learning, remembering, and understanding. In J. Flavell & E. Markman (Eds.), Handbook of child psychology: Vol. 3. Cognitive development (pp. 77 - 166). New York: Wiley.

Brown, R. (1989). Testing and thoughtfulness. Educational Leadership, 46, 31-33.

Corno, L. (1986). The metacognitive control components of self-regulated learning. Contemporary Educational Psychology, 11, 333-346.

Corno, L., & Mandinach, E. (1983). The role of cognitive engagement in classroom learning and motivation. Educational Psychologist, 18, 88-100.

Duffy, T. M. & Bednar, A. K. (1991). Attempting to come to grips with alternative perspectives. Educational Technology, 31, (September), 12-15.

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. Journal of Personality and Social Psychology, 54(1), 5-12.

Gabbert, B., Johnson, D. W., & Johnson, R. T. (1987). Cooperative learning group-to-individual transfer, process gain, and the acquisition of cognitive reasoning strategies. The Journal of Psychology, 120(3), 265-278.

Greene, B. A., & Miller, R. B. (1993, April). Influences on course performance: Goals, perceived ability, and self-regulation. Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA.

Griffin, M. M., Quinn, G. P., McCown, R. R., & Driscoll, M. P. (1994, April). Learning through testing: An investigation of cooperative assessment. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Johnson, R. T., Johnson, D. W., & Smith, K. A. (1990). Cooperative learning: An active learning strategy for the college classroom. Baylor Education, 15(2), 11-16.

Kember, D., & Murphy, D. (1990). Alternative new directions for instructional design. Educational Technology, 30(8), 42-47.

Lambiotte, J. G., Dansereau, D. F., Rocklin, T. R., Fletcher, B., Hythecker, V. I., Larson, C. O., & O'Donnell, A. M. (1987). Cooperative learning and test taking: Transfer of skills. Contemporary Educational Psychology, 12, 52-61.

Nystrand, N. (1986). The structure of written discourse: Studies of reciprocity between readers and writers. New York: Academic Press.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. Journal of Educational Psychology, 82(1), 33-40.

Sarason, I., Sarason, B., & Linder, K. (1983). Assessment and experimentally provided social support. Seattle: University of Washington Press.

Shavelson, R. J., Baxter, G. P., & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. Educational Researcher, 21(4), 22-27.

Singer, R. S. (1990). The effects of collaborative testing on the test scores and the classroom attitudes of junior high school pre-algebra students. ERIC Document TM016072.

Shepard, L. A. (1989). Why we need better assessments. Educational Leadership, 46, April, 4-9.

Slavin, R. E. (1991). Synthesis of research on cooperative learning. Educational Leadership, 48, (February), 71-82.

Vygotsky, L. (1978). Mind and society. Cambridge, MA: Harvard University Press.

Webb, N.M. (1993). Collaborative group versus individual assessment in mathematics: Processes and outcomes. Educational Assessment, 1(2), 131-152.

Wiggins, G. (1989). Teaching to the (authentic) test. Educational Leadership, 46, (April), 41-47.

Wynne, E. A. (1983). Teaching about cooperation. Phi Delta Kappan, 64, 497-499.

Appendix A

Means, Modes, and Standard Deviations for the Group Assessment Survey

Item		Mean	Mode	SD
1.	To what extent was each group member prepared for the exams? (unprepared to very well prepared)	6.04	6	0.94
2.	To what extent did each group member participate and offer input as you completed the group exams? (did not participate or offer input to participated and offered extensive input)	6.47	7	0.94
3.	My group members seemed to study less and rely on the group's success to raise their exam grades.	1.70*	1	1.26
4.	In general, our group discussed and debated the answers to most of the test items.	5.78*	7	1.68
5.	What effect did taking the group exam have on your study habits? (I studied much less than usual to I studied much more than usual)	4.77	4	0.97
6.	What effect did the difficulty level of the exams have on your study habits? (I studied much less than usual to I studied much more than usual)	5.30	6	1.08
7.	Taking the tests as part of a group was beneficial to my grade.	6.10*	7	1.19
8.	Taking the tests as part of a group was beneficial to my understanding of required course concepts.	6.42*	7	0.77
9.	In general, our group did not debate the responses to most items, but relied upon one or two members to provide the answers.	2.33*	1	1.93
10.	The group as a whole worked together to complete all of the essay items.	4.94*	6, 7	1.91
11.	As I completed the group exam, I gained a better understanding of content I missed when I took the test individually.	6.50*	7	0.83

12.	As I completed the group exam, I was more concerned with determining which items I answered correctly on my own than with learning from the group.	3.00*	4	1.41

*All items used a 7-point Likert scale, with higher numbers indicating more positive responses with the exception of items 3, 9, and 12 in which case the lower numbers indicate a more positive response. End points of the Likert scale for all items (unless otherwise indicated following the item) were "Strongly disagree" (1) to "Strongly agree" (7).

Appendix B

Interview Questions and Response Summaries

I. Affective/Process Outcomes

A. Goal

Q: What was the purpose of the peer-group test?

1. Problem solving
   - to explain why answers are correct or incorrect (clear up confusing points or wrong answers)
   - discuss reasoning behind answers
2. Multiple views are highlighted
   - to see others' perspectives, others' thinking
3. Benefits of group discussion
   - discussion should improve learning, memory, understanding

B. Social/Behavioral Interactions of Group Members

Q: Did everybody in the group do about the same amount of talking or did one particular person dominate the discussion? Was it a pleasant experience?

1. Amount of talking
   - students who knew more or had more experience talked more, but most felt that all students had equal power
2. Pleasant experience
   - pleasant experience to discuss the test with peers

Q: How did you respond to each other's comments during the group testing session? Can you give me an example of your response to someone in the group?

Students typically
- engaged in positive exchanges (equity, respect, fairness, acceptance, no hostility)
- tried to understand the reason behind peers' responses to items
- tried to evaluate the best answer

C. Self-efficacy

Q: Was the group test a difficult challenge for you and for your group? Why or why not?

- students apparently did not feel challenged except when a lot of disagreement occurs, then they feel challenged to resolve the disagreement

II. Cognitive/Metacognitive Outcomes

A. Planning

Q: Did you do anything special to prepare for the peer-group testing activity outside of class? If so, what?

1. Study strategies
   - Group assessment did not change students' typical methods (strategies) for test preparation. Rather, they tried to apply their usual study techniques to "fit" the group discussion activity.
2. Level of preparation
   - some felt more responsible to be more prepared than usual because they were concerned about peer pressure

Q: What was your goal during the activity? What did you hope to accomplish?

1. External goal
   - All students mentioned increasing grades, number of points, or score on test. This seemed to be their primary goal.
2. Internal goal
   - 7 of 8 students also mentioned that they wanted to learn, to understand, to get other perspectives on the material, or to clarify misunderstandings

B. Monitoring and Regulating

Q: What was the most difficult part of the group test?

1. Disparity of answers
   - difficulty occurred especially when there was great disparity in group answers
2. Consensus-seeking
   - trying to get everyone to agree which was difficult because either
     1) no one is sure of the content (lack of knowledge)
     2) each person is more likely to believe their answer is correct
     3) more persuasion had to be performed, which requires more effort

Q: Did you spend you time on some particular items more than others? If so, why?

1. Multiple choice items
   - particular items created more disagreement, taking more time to work through discussion and reach consensus
2. Essay
   - required students to first decide who responded to which item, to discuss individual answers, and to choose or construct best group response

Q: Did you ever go back to items you already answered? Why?

-students paid attention to and utilized test taking strategies to regulate their thinking, such as noting difficulty of item, elapsed time, illumination of previous response by later discussion

Q: What did you do when you didn't know an answer or didn't know if your answer was correct? Various strategies used

- went with best explanation or example, deciding this person must know answer
- temporarily skip the item and come back to it
- went with majority opinion
- went with "expert" opinion (rarely done)
- guessed (rarely done)

Q: When you were taking the group test, how did you decide on the best answers to the multiple choice items? Give an example.

•First, groups would ask for each person's response to an item. Procedures varied based on consensus or lack of consensus.

1. Consensus
   - If all agreed, then go on to next item with little or no discussion
2. Lack of consensus
   - some disagreement, but majority agree then groups discussed a little bit but often opted for majority view
   - major disagreement, with equal representation of two responses lead to much discussion and a search for the best explanation or justification, followed by a decision on correct response

Q: When you were taking the group test, how did you decide on the answers to the essay? Give an example.

•First students determined which essay question was answered most frequently.

1. All or majority answered same question
   - each reported and explained their answer, then group discussed the answers and selected the best response to use as the group response
2. If different questions were answered
   - group decided on one to respond to, but some people were left out of the response generation

C. Evaluating Activity

Q: What did you learn in the group testing process?

Process helped to clarify information

- generated more and diverse examples of concepts from different perspectives
- remembered teacher's or textbook's examples (review)
- clarified relationships among concepts
- applied concepts rather than merely memorizing definitions